Analysis of expressed sequence tags from the four main developmental stages of *Trypanosoma congolense*☆

Jared R. Helm [a,1], Christiane Hertz-Fowler [b,1], Martin Aslett [b], Matthew Berriman [b], Mandy Sanders [b], Michael A. Quail [b], Marcelo B. Soares [c,2], Maria F. Bonaldo [c,2], Tatsuya Sakurai [d], Noboru Inoue [d], John E. Donelson [a,*]

[a] *Department of Biochemistry, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA*
[b] *Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK*
[c] *Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA*
[d] *National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan*

ARTICLE INFO

ABSTRACT

Trypanosoma congolense is one of the most economically important pathogens of livestock in Africa. Culture-derived parasites of each of the three main insect stages of the *T. congolense* life cycle, *i.e.*, the procyclic, epimastigote and metacyclic stages, and bloodstream stage parasites isolated from infected mice, were used to construct stage-specific cDNA libraries and expressed sequence tags (ESTs or cDNA clones) in each library were sequenced. Thirteen EST clusters encoding different variant surface glycoproteins (VSGs) were detected in the metacyclic library and 26 VSG EST clusters were found in the bloodstream library, 6 of which are shared by the metacyclic library. Rare VSG ESTs are present in the epimastigote library, and none were detected in the procyclic library. ESTs encoding enzymes that catalyze oxidative phosphorylation and amino acid metabolism are about twice as abundant in the procyclic and epimastigote stages as in the metacyclic and bloodstream stages. In contrast, ESTs encoding enzymes involved in glycolysis, the citric acid cycle and nucleotide metabolism are about the same in all four developmental stages. Cysteine proteases, kinases and phosphatases are the most abundant enzyme groups represented by the ESTs. All four libraries contain *T. congolense*-specific expressed sequences not present in the *Trypanosoma brucei* and *Trypanosoma cruzi* genomes. Normalized cDNA libraries were constructed from the metacyclic and bloodstream stages, and found to be further enriched for *T. congolense*-specific ESTs. Given that cultured *T. congolense* offers an experimental advantage over other African trypanosome species, these ESTs provide a basis for further investigation of the molecular properties of these four developmental stages, especially the epimastigote and metacyclic stages for which it is difficult to obtain large quantities of organisms. The *T. congolense* EST databases are available at: http://www.sanger.ac.uk/Projects/T_congolense/EST_index.shtml.

1. Introduction

The *Trypanozoon* subgenus of African trypanosomes are protozoan parasites of major socio-economic impact in sub-Saharan Africa. Included in this group are the three "*brucei*" subspecies, which are transmitted by tsetse flies [1]. Two of the subspecies, *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense*, cause human sleeping sickness in East Africa and in West/Central Africa, respectively, and the third, *T. b. brucei*, is responsible for a small amount of livestock disease in Africa but does not survive in human serum [2,3]. Most laboratory-based research on African trypanosomes during the past four decades has been conducted on *T. b. brucei* because it does not pose as much of a laboratory safety concern as the two human-infective subspecies and two developmental stages of its life cycle, the procyclic form in the insect and

the bloodstream form in the mammal, can be readily maintained in culture. The 35-Mb haploid genomic DNA sequence of a specific *T. b. brucei* clone has been determined [4].

The most important trypanosome species for the African live-stock industry, however, is *Trypanosoma congolense* [5], which belongs to the *Nannomonas* subgenus and is also transmitted by tsetse flies. *T. congolense* causes nagana, a chronic wasting disease of cattle that is characterized by anemia, weight loss and immuno-suppression and is typically fatal if untreated. It has the largest impact by far of any trypanosome species on domestic livestock in Africa. Economic losses in Africa due to *T. congolense* infections of livestock have been conservatively estimated as US$1.3 billion annually [6]. *T. congolense* can be subdivided further on the basis of isoenzyme analysis, repetitive DNA detection and geographic location into three subgroups: Savannah, Forest, and Kilifi (coastal region). The Savannah group is the most prevalent (90% of stocks) followed by the Forest (9%) and Kilifi (1%) groups [7]. A genome sequencing project on *T. congolense* Savannah clone IL3000 [8] is underway [http://www.sanger.ac.uk/Projects/T_congolense/]. The work described here was conducted with this same *T. congolense* clone IL3000.

T. congolense and the *T. brucei* subspecies have similar, but non-identical, life cycles *in vivo*. Bloodstream *T. brucei* occurs as dividing long slender, intermediate forms as well as a non-dividing stumpy form, whereas bloodstream *T. congolense* has the corresponding dividing forms but the stumpy form is less obvious [9]. In the nutrient-rich, 37 °C bloodstream the main challenge for both species is to avoid the mammalian immune system. Their bloodstream forms possess about 10^7 copies of a variant surface glycoprotein (VSG) (~5% of the total cellular protein) [10], and periodically switch from one VSG to another in a successful effort to avoid the mammalian immune responses [11]. When bloodstream trypanosomes are consumed by a tsetse fly during its blood meal, they enter the insect gut and differentiate to the procyclic form, which must adapt to cooler and variable temperatures, new nutrient sources, hostile digestive enzymes and an insect immune system (see [12] for a review of the interactions between African trypanosomes and tsetse flies). This differentiation from bloodstream to procyclic form is associated with (i) morphological changes, (ii) a switch from glycolysis to oxidative phosphorylation for energy metabolism, and (iii) the replacement of the VSG with procyclic-specific surface proteins [13]. The procyclic form and subsequent epimastigote form *T. congolense* contain on their surface a procyclin protein consisting almost exclusively of 13 heptapeptide repeats (EPGENGT) [14], as well as an unrelated invariant protein, GARP (glutamic acid/alanine-rich protein) [15–17]. A *T. congolense* "*congolense* epimastigote-specific protein" (CESP) has recently been reported [18]. The function(s) of these procyclic- and epimastigote-specific proteins have not been explicitly determined [16,18,19], but they may protect against proteases and other enzymes in the tsetse fly midgut and/or they may play a role in determining the sites of subsequent trypanosome development in the fly [16,19].

Studies tracking the progress of *T. congolense* through flies are much less extensive than for *T. brucei*. The development from procyclic to epimastigote form appears to be similar to *T. brucei*, and the most obvious difference is that for unknown reasons *T. congolense* epimastigotes are directed, not to the fly's salivary glands as are *T. brucei* epimastigotes, but to its proboscis and mouth parts where the epimastigotes differentiate into non-dividing, infective metacyclic forms (reviewed in [12]). Similar to *T. brucei*, immunofluorescence studies using monoclonal antibodies suggest these individual *T. congolense* metacyclic organisms possess on their surface one of about 12 different metacyclic VSGs [20,21], instead of the repertoire of as many as 1000 or more VSGs available to bloodstream *T. brucei* [4]. Metacyclic trypanosomes are inoculated during the fly bite into the host, where parasites multiply at the site of the bite and

for 5–9 days continue to express metacyclic VSGs [20–22]. The life cycle is complete when the parasites begin to express bloodstream VSGs on their surface and invade the bloodstream and lymphatic system.

We took advantage of the ability to culture the three insect stages of *T. congolense in vitro*, *i.e.*, the culture-derived procyclic (PCF), epimastigote (EMF) and metacyclic (MCF) forms [23,24], and to obtain the bloodstream form (BSF) from laboratory mice infected with the cultured MCF organisms, so that cDNA libraries could be prepared from each of these four developmental stages. We also prepared normalized cDNA libraries from two forms, MCF and BSF. We report here an analysis of expressed sequence tags (ESTs) in each of these six libraries.

2. Methods and materials

2.1. Trypanosomes and culturing conditions

The bloodstream *T. congolense* (Savannah group) IL3000 clone [8], previously obtained from the International Livestock Research Institute (ILRI; Nairobi, Kenya), was maintained at the National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan. All animal experiments were conducted in accordance with the standards relating to the Care and Management of Experimental Animals at Obihiro University of Agriculture and Veterinary Medicine (No. 14-69).

A frozen stabilate of the BSF IL3000 clone was thawed and aliquots inoculated intraperitoneally into female, 8-week-old BALB/c mice obtained from CLEA Japan Inc. (Tokyo). Infections were monitored via tail nicks and at the first peak of parasitemia the mice were bled by cardiac puncture and blood collected in heparin. Trypanosomes were purified from whole blood on DE-52 anion exchange column chromatography (Whatman plc., Middlesex, UK) [25]. *In vitro* cultures of PCF, EMF and MCF cells derived from these BSF IL3000 cells were obtained following the methods of Hirumi and Hirumi [23]. Briefly, the concentration of the BSF cells was adjusted to 3×10^6 cells/ml in Eagle's minimum essential medium supplemented with 20% fetal bovine serum, 2 mM L-glutamine and 10 mM L-proline. This parasite suspension was transferred to 25 cm^2 culture flasks (10 ml/flask) and the flasks kept in an incubator at 27 °C for a week, during which the BSF cells differentiated to PCF cells. The PCF cells were maintained and amplified by changing medium and preparing subcultures. After 1–2 months, adherent EMF parasites began to appear as small clusters on the bottom surface of the flasks. The clusters increased in size and number, and finally covered the whole bottom surface. When the culture was primarily adherent EMFs on the flasks' bottom surface, the non-adherent MCF parasites began to appear in the supernatant of the confluent EMF cultures [26].

For isolation of the different life-cycle forms, the PCF cells were collected from the culture supernatant by centrifugation at 1500 × g for 10 min at 4 °C. They were washed 3 times with phosphate-buffered saline (PBS). EMF cells were collected from confluent culture flasks by scraping the adherent cells from the bottom of the culture flasks. Briefly, the flasks were washed gently three times with 10 ml PBS to remove any remaining PCF cells and newly appearing MCF cells. The EMF cells were removed by scraping from the bottom surface of the flasks and suspending in PBS. The EMF cells were washed three times with PBS by centrifugation at 1500 × g for 10 min at 4 °C. The MCF cells (which possess surface VSGs) were separated from the EMF and PCF cells (which do not have surface VSGs) in a mixed culture of MCF, EMF and PCF cells by passing the culture through a DE-52 anion exchange column [23,25]. MF cells passed through the column, while PCF and EMF cells were retained. These *in vitro*-generated MCF cells were used

to inoculate BALB/c mice (1×10^5 cells/animal) intraperitoneally. Seven days after infection, the blood was collected by cardiac puncture and the new BMF cells purified on a DE-52 column [25].

2.2. EST library construction

Total RNA from PCF, EMF and MCF cells and from BSF cells obtained from mice 7 days after inoculation with the MCF cells was isolated either by the method of Chomczynski and Sacchi [27] or by using a TRIzol®Regent kit (Invitrogen, Carlsbad, CA, USA) according to manufacture's instructions. Poly(A)-containing mRNA was prepared from the total RNA isolations using the polyATtract® mRNA Isolation System (Promega, Madison, WI, USA). Six EST libraries were prepared: four from the mRNAs of each of the four cell types and two additional normalized EST libraries from the MCF and BSF mRNAs. For the library constructions, the first-strand cDNA synthesis by reverse transcriptase was primed with a *Not*I-tag-oligo-(dT)$_{18}$. The tag is a unique sequence of 10 nucleotides that serves as an identifier for that library. The resulting DNA/RNA hybrid was treated with RNase H and used as a template for second-strand synthesis by *E. coli* DNA polymerase I. *Eco*RI adaptors were added to the ends, and the double-stranded cDNAs digested with *Not*I and size selected. The resulting molecules were cloned into the *Eco*RI and *Not*I sites of the phagemid vector pT7T3PAC. Aliquots of the MCF and BCF libraries were then subjected to one round of normalization performed according to 'method 4' [28,29]. This procedure is based on the hybridization of PCR-amplified cDNA inserts of a library with the library itself in the form of single-stranded circles. Following hybridization to a relatively low renaturation Cot of 5–10, the remaining single-stranded circles (normalized library) were purified over hydroxyapatite, converted to double-stranded circles by primer extension and electroporated into bacteria [29]. The titres of the libraries ranged from 2×10^6 to 5×10^7 clones/ml.

2.3. Sequencing and bioinformatics

Randomly selected bacterial clones in each of the six EST libraries were subjected to DNA sequencing (Table 1). Both sides of the inserts were sequenced with T3 and T7 primers using the ABI™ Big Dye Terminator Cycle Sequencing kits. The sequences were clipped for vector sequence using Cross Match (P. Green, unpublished) and for quality using Phred, a base-calling program for automated sequencer traces that assigns an error probability to each called base (P. Green, unpublished). Sequence reads were assembled into clusters with Phrap (P. Green, unpublished; www.phrap.org). Where necessary, clusters were inspected and manually finished using the Gap4 software [30].

Each EST cluster was analyzed for similarities to published protein sequences using BLASTX against the UniprotKB database [31] and using a blast bit score of greater than 50 as a cut-off. VSG coding

sequences were manually annotated in Artemis [32], according to previously published standards [4]. Alignments of VSG protein sequences were produced using the clustalw software with default parameters (http://www.ebi.ac.uk/Tools/clustalw2/index.html) [33]. The EST clusters are accessible via the *T. congolense* project pages at the Wellcome Trust Sanger Institute website (http://www.sanger.ac.uk/Projects/T_congolense/EST_index.shtml). They have also been deposited in the EMBL database under the following accession numbers: FN263376–FN292969.

3. Results and discussion

A total of 20,449 clones from the six different libraries were sequenced, corresponding to between 2900 and 3800 EST clones for each library (Table 1). In all six libraries one end of about 30% of the ESTs had at least a few 3′ nucleotides of the 39-nucleotide spliced leader found at the 5′ ends of all mature trypanosome mRNAs [11], reflecting the fact that first-strand cDNA synthesis had proceeded from the 3′ poly(A) of the template mRNA molecule to its 5′ end. ESTs lacking the 5′ spliced leader sequence were usually considered to be partial length, an assumption borne out for many ESTs with homology to genes with coding sequences of 2 kb or larger.

3.1. Assessment of in vitro differentiation

The presence, absence and relative abundances of transcripts encoding known surface proteins were used as a measure of success of the *in vitro* differentiation process and subsequent cell separation procedures. Given that PCF, EMF and MCF cells were all derived from the same continuous culture (see Section 2.1), it was important to establish the purity of the cell population used as source material. The presence of VSGs has typically served as a molecular marker for the metacyclic and bloodstream stages of African trypanosomes with procyclins as markers for the procyclic stage. Markers for the epimastigote stage are less extensively characterized, but in *T. congolense* the recently identified "*congolense* epimastigote-specific protein" (CESP) serves that role [18]. In addition, GARP is thought to be weakly expressed in *T. congolense* procyclic cells and abundant in epimastigotes [18,34]. By the criteria of VSGs and procyclins, the MCF cells were well separated from the EMF cells with 5.8% of the ESTs (224/3843) in the MCF library encoding VSGs and none of the MCF ESTs encoding procyclin, GARP or CESP (Fig. 1). Near identical results were obtained for the BSF library. These VSG data are consistent with previous estimates that VSGs constitutes about 5% of the total cellular protein in these two developmental stages [10].

No VSG ESTs were detected in the PCF library (Fig. 1), consistent with the interpretation that the PCF cells had completely differentiated from the mouse-blood-derived BSF cells used to initiate the PCF culture. Somewhat unexpectedly, however, in the EMF library 0.3% of the ESTs (12 out of 3792) were found to have one of five dif-

Table 1
Summary of ESTs sequenced.

EST library	Number of EST clones sequenced	% identified ESTs[a]	% hypothetical ESTs[b]	% "no nit" ESTs[c]
Procyclic (PCF)	3,423	47%	26%	27%
Epimastigote (EMF)	3,792	41%	27%	32%
Metacyclic (MCF)	3,843	34%	21%	45%
Bloodstream (BSF)	3,523	39%	28%	33%
norm. Metacyclic (normBSF)	2,892	25%	31%	42%
norm. Bloodstream (normBSF)	2,976	30%	34%	36%
Total	20,449			

[a] % of ESTs with homology to genes whose encoded proteins have known functions in other organisms. These ESTs do not include the "hypothetical" ESTs or ESTs with no match >50.

[b] % of ESTs with homology to "hypothetical" genes in other organisms whose potentially encoded proteins have no known function.

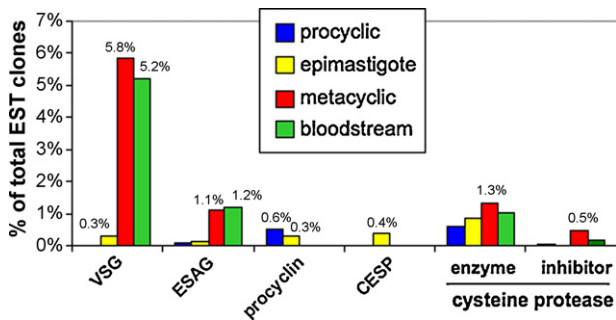[c] % of ESTs with no match >50 to known sequences in other organisms.

Fig. 1. Bar graph showing the abundances of the indicated EST clusters in each of the four regular (non-normalized) cDNA libraries. VSG, variant specific glycoprotein; ESAG, expression-site associated gene; CESP, *congolense* epimastigote-specific protein.

ferent VSG coding sequences, yet EMFs are thought not to express VSGs. This finding is made more unexpected by the fact that these five VSG EST clusters in the EMF library are different than any of the VSG EST clusters detected in the MCF library, indicating they were not due to contamination of the harvested EMF cells with MCF cells (see below). Another unexpected result was that in the PCF and EMF libraries the procyclin ESTs constituted only 0.6% and 0.3% of the ESTs, respectively (Fig. 1). Since procyclin molecules replace the VSG molecules on the trypanosome surface during differentiation from BSF to PCF, it might be expected that the procyclin mRNA levels in PCF cells would be closer to the 5% level observed for VSG

mRNAs in MCF and BSF cells. GARP ESTs were also less frequent than expected (not shown). It is worth noting that a previous study reported that *T. congolense* epimastigote cells in flies 14 days post-infection lacked detectable GARP [34], so the relative abundance of GARP during the different *T. congolense* stages in the fly, and perhaps in different *T. congolense* isolates, remains uncertain. ESTs encoding CESP constituted 0.4% of the EMF library (17/3792) and were not detected in the other libraries (Fig. 1), consistent with the recent finding that CESP is specific to EMFs [18].

Finally, ESTs with homologies to *T. brucei* VSG expression-associated site genes (ESAGs) were much more abundant in the MCF/BSF libraries compared to the PCF/EMF libraries (Fig. 1). Although no sequences of VSG gene expression sites in *T. congolense* have been published, the presence of these ESAG-like ESTs in the same libraries that contain VSG ESTs is also consistent with good separation of the four different developmental stages prior to cDNA library construction.

3.2. VSG gene expression in MCF and BSF cells

The sequences of several *T. congolense* VSG genes have been previously reported, including two that encode VSGs expressed at the metacyclic stage (mVSGs) [35–37]. Nascent VSGs have different amino acid sequences, but share a high conservation of cysteine positions, whose pattern is different in VSGs isolated from *T. congolense* and *T. brucei*. Fig. 2A shows the relative positions of eight highly conserved cysteines in the known *T. congolense* VSGs [37]. Based on these properties, the VSG ESTs in the MCF library
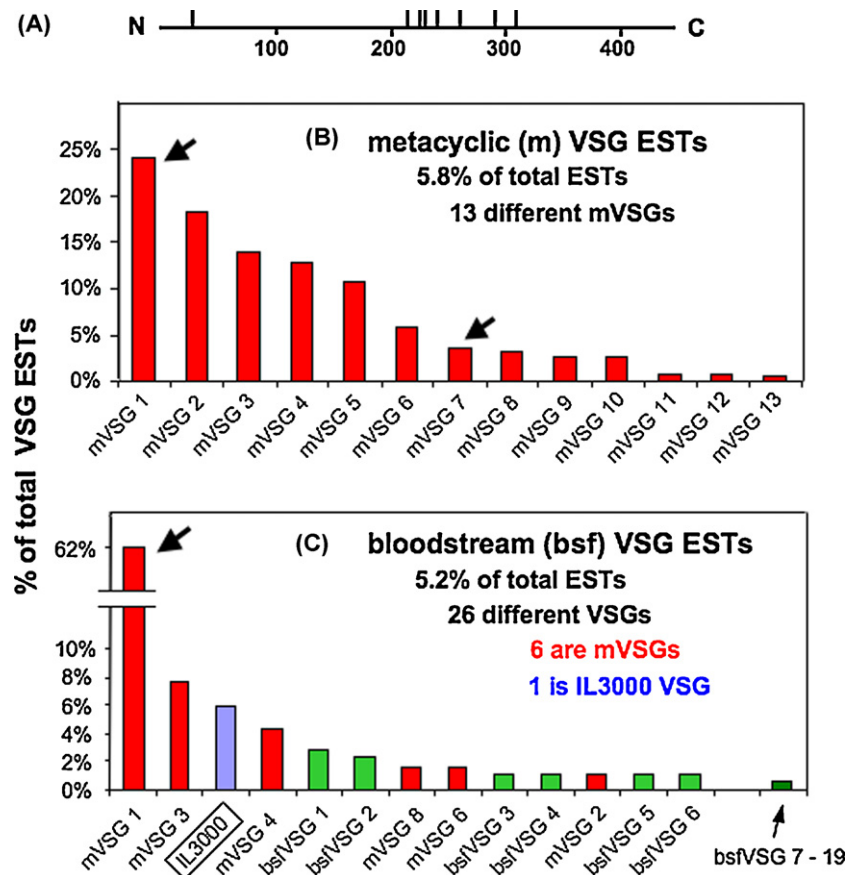

Fig. 2. (A) Line diagram depicting the amino acid sequences of nascent *Trypanosoma congolense* VSGs with the N-terminus on the left and C-terminus on the right. Numbers are approximate amino acid positions and vertical bars indicate conserved cysteines (redrawn from [37]). (B) Bar graph showing the relative abundances of the 13 mVSG EST clusters containing 224 mVSG EST clones identified in the MCF EST library. Arrows indicate two previously identified *T. congolense* mVSGs [35]. (C) Bar graph showing relative abundances of the 26 VSG EST clusters containing 184 VSG clones identified in the BSF EST library. Note that 6 of these 26 clusters encode mVSGs, including the most abundant cluster, which constitutes 62% of the 184 VSG clones (see text). IL3000 VSG was expressed by bloodstream *T. congolense* clone IL3000 used to generate the four developmental stages. Each of the bsfVSG groups 7–19 has only a single VSG EST.

(5.8% of the MCF ESTs) were found to encode 13 different mVSGs (Fig. 2B), consistent with previous estimates of 12–15 mVSG types [20,21,38,39]. Similarly, the VSG ESTs detected in the BSF library (5.2% of the total) encode 26 different VSGs (bsfVSGs) (Fig. 2C). The relative abundances of the 13 mVSG EST clusters in the MCF library ranged from 24% of the mVSG ESTs for mVSG1 to 1% for mVSG13 (Fig. 2B). The mVSG1 and mVSG7 coding sequences (indicated by arrows in Fig. 2B) are the same two mVSG sequences of *T. congolense* IL3000 identified earlier [35]. The five most abundant mVSG ESTs (mVSG1–mVSG5) account for 80% of the mVSG ESTs. The relative abundances of the these IL3000 mVSG ESTs are similar to the relative abundances of 12 mVSG proteins previously reported in another *T. congolense* stock, TREU 1290 [21].

The BSF library was constructed from RNA of BSF trypanosomes collected from mice 7 days after the mice had been inoculated with MCF organisms (see Section 2). Of the 26 different VSG sequences detected in this BSF library, 6 are mVSG sequences (red bars in Fig. 2C). The most abundant mVSG EST, mVSG1 (24% of the mVSG ESTs), is even more abundant in the BSF library (62%). This continuation of mVSG expression 7 days after inoculation is consistent with the observation that *T. congolense* mVSGs can continue to be expressed by the rapidly dividing BSF cells in the blood for as long as 9 days after infection with non-dividing MCF cells [21]. However, it was unexpected that the mVSG1 EST would be 2.5 times more abundant in the library of these 7-day BSF cells than in MCF library. This increase in abundance is in contrast to the decreased abundance observed with the other mVSGs. The simplest of several possible explanations for these individual variations in continued mVSG expression early in the bloodstream infection is that some telomere-linked mVSG expression sites [20] are more stably transcribed than others after differentiation from MCF to BSF, but much remains to be learned about the on-off regulation of mVSG gene expression sites [11,20].

The most abundant non-mVSG EST in the BSF library (the blue bar in Fig. 2C) encodes the IL3000 VSG, *i.e.*, the VSG expressed by the BSF *T. congolense* clone [8] used to establish the initial *in vitro* procyclic cell culture. This result is consistent with earlier observations that the bsfVSG expressed by BSF trypanosomes ingested by a tsetse fly is often among the first bsfVSGs to be expressed when metacyclic trypanosomes from that fly infect a new animal [22]. In this case, the IL3000 VSG EST is 6% of the total VSG ESTs in the BSF library, whereas the remaining 19 different bsfVSG EST clusters detected are each only 1–3% of the total. Thus, aside from the IL3000 VSG, no VSG emerged as a dominant bsfVSG to be expressed immediately after the switch from expression of mVSGs to bsfVSGs.

3.3. VSG ESTs in the EMF library

Since VSGs are not thought to be present on epimastigote cells, the five VSG EST clusters detected in the EMF library mentioned above were examined more closely. The number of clones in each of the five clusters was 5, 3, 2, 1 and 1, respectively, for a total of 12 VSG ESTs detected in the EMF library (0.3% of the total EMF ESTs examined). None of these five VSG EST clusters is the same as any of the 13 mVSG EST clusters detected in the MCF library. Four of the five are also different than the VSG EST sequences detected in the BSF library. None of these four has an open translation reading frame for a functional VSG, suggesting they may have been derived from a pseudo-gene (95% of the approximately 1000 VSG genes in the *T. brucei* genome are pseudo-genes [4]). The fifth EMF VSG sequence is encoded by a single read of 560 bp in a single clone that is identical to a portion of the 1.5-kb bsfVSG2 EST sequence in the BSF library (Fig. 2C). No further information on this sequence in EMF cells was obtained. Thus, on the basis of these five EST clusters, it appears that a low level of VSG gene transcription occurs in EMF cells, but at least some of these transcripts do not encode
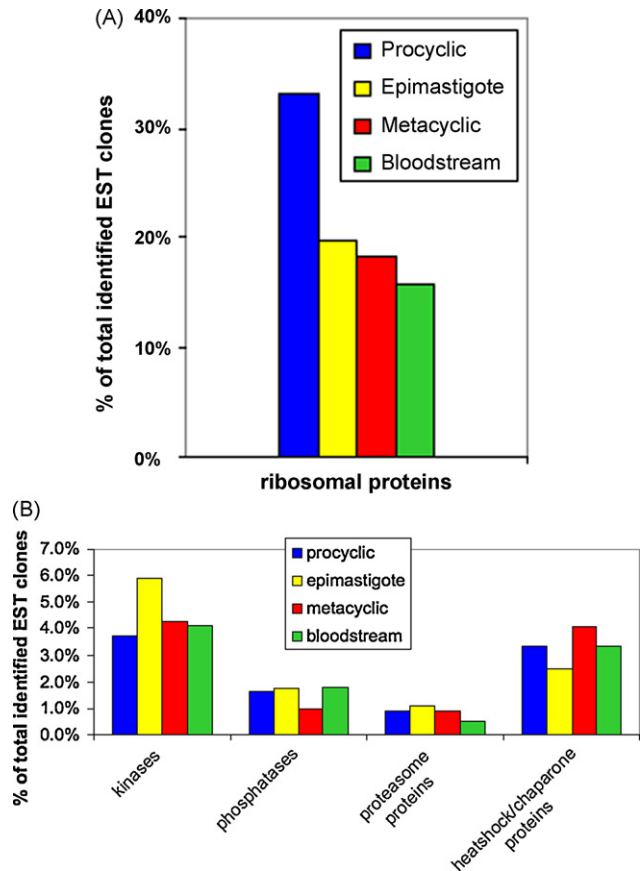

Fig. 3. Bar graphs showing the relative abundances of ESTs encoding (A) ribosomal proteins and (B) the indicated protein families in each of the four regular libraries.

functional VSGs. Other possibilities are that these transcripts are not translated, their protein products are rapidly degraded, or only an insignificant amount of VSG or truncated VSG protein is made from their rare mRNAs in EMF cells.

3.4. Other abundant transcripts

Of the four regular (non-normalized) libraries, the MCF library had the lowest percent (34%) of ESTs with homology to genes encoding proteins of known functions in other organisms, and the PCF library had the highest percent (47%). Likewise, the MCF library had a slightly lower percent (21%) of ESTs with homology to genes encoding proteins of unknown function than did the other libraries (26–28%). As expected [40], most homologies were to genes from the published genomes of *T. brucei* [4], *T. cruzi* [41] and *Leishmania major* [42], and with some exceptions, the homologies were greatest to predicted proteins in *T. brucei*, followed by *T. cruzi*, followed by *L. major*. As a corollary, of these four regular libraries, the MCF library had the highest percent of 'no hit' ESTs bearing no matches to known sequences (45% versus 27–33% in the other libraries), an observation discussed below. In the two normalized libraries (normMCF and normBSF), the percent of 'identified' ESTs was somewhat lower (25–30%) than in the corresponding regular libraries (34–39%), which is consistent with the expected removal of the most abundant EST sequences by the normalization process and enhancement of less abundantly expressed sequences.

As with other organisms, the most abundant ESTs, by far, encode ribosomal proteins (Fig. 3A). Ribosomal-protein ESTs were 33% of the PCF library and 15–19% of the other three libraries, suggesting the possibility that procyclic cells have more ribosomes and translation capacity than the other three cell types. Comparisons of the
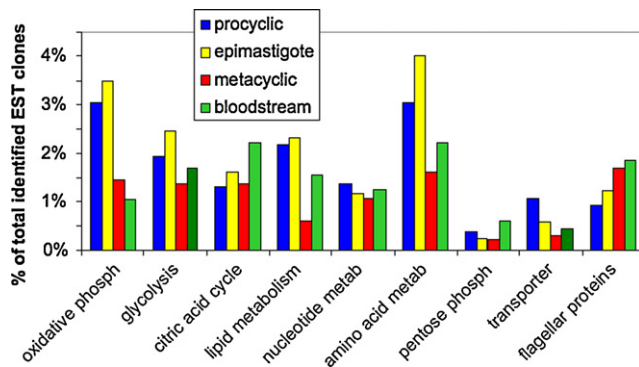
Fig. 4. Bar graph showing the relative abundances of EST clusters encoding enzymes of the indicated metabolic pathways, transporters or flagellar proteins.

ESTs encoding regulatory proteins, such as kinases, phosphatases, proteasome proteins and heat shock/chaperone proteins, revealed few differences among the four developmental stages (Fig. 3B).

3.4.1. Metabolic pathways

Fig. 4 shows the relative abundance of ESTs encoding proteins that catalyze the indicated metabolic pathways or contribute to nutrient transport and the flagellar cytoskeleton. Procyclic cells predominately utilize mitochondrial oxidative phosphorylation as an energy source, whereas bloodstream cells have only a vestige of a mitochondrion and primarily use glycolysis as an anaerobic energy source (reviewed in [43,44]). The relative EST abundance is consistent with these metabolic properties. ESTs encoding proteins involved in oxidative phosphorylation are 2–3 times more abundant in PCF and EMF cells (3–3.5% of total) than in BSF cells (1–1.4% of total) (Fig. 4). MCF cells have only slightly more of these oxidative phosphorylation ESTs (1.4% of total) than do BSF cells (1%), indicating they pre-adapt to the bloodstream environment by both acquiring a VSG coat and down-regulating expression of genes encoding oxidative phosphorylation proteins. The abundance of ESTs encoding glycolytic enzymes, on the other hand, varies less among the four developmental stages, ranging from 2.4% in EMF cells to 1.4% in MCF cells, with PCF and BSF having intermediate values of 1.9–1.7%. BSF cells have only a slightly higher percentage of ESTs encoding citric-acid-cycle enzymes (2.2%) than do the other three stages (1.3–1.6%), reflecting the fact that the citric acid cycle is used by the insect stages of African trypanosomes for purposes other than degradation of mitochondrial substrates [43–45].

PCF and EMF cells have a higher percent of ESTs encoding enzymes of amino acid metabolism than do MCF and BSF cells, with EMF cells having the most with 4% of their ESTs devoted to this category, underscoring the known importance of amino acid metabolism in the insect stages [43,44]. The percent of ESTs involved in lipid metabolism (fatty acids and sterols) is about 2.3% in PCF and EMF, whereas MCF cells have only about 1/4th of that (0.6%), for reasons that are unclear. The percent of ESTs encoding enzymes of nucleotide metabolism is the about same in all four stages (1.1–1.4%), whereas membrane transporter ESTs are more abundant in the PCF (1.2%) than in the other stages (0.7–0.4%).

3.4.2. Cytoskeleton

Two curiosities exist with the ESTs encoding β- and α-tubulin (Fig. 5). First, in three of the four stages (PCF, EMF and BSF), the percent of β-tubulin ESTs is 2.0–2.5 times greater than the α-tubulin ESTs, despite the fact that these two related proteins are equivalently abundant as the heterodimeric building block of microtubules [46]. In contrast, in MCF cells the ratio of ESTs encoding β-tubulin versus α-tubulin is closer to the expected ratio of 1.0 (1.2%). Secondly, the percent of β- and α-tubulin ESTs is
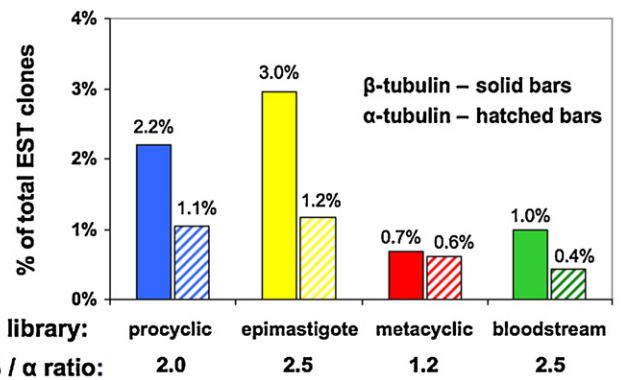

Fig. 5. Bar graph comparing β- and α-tubulin EST abundances in the indicated libraries.

twice as high in PCF/EMF cells as in MCF/BSF cells. These anomalies may be consistent with studies in other eukaryotes showing that expression of β- and α-tubulin proteins can be influenced by many post-transcriptional events, including an autoregulatory one in which the nascent N-terminal tetrapeptide emerging from the ribosome can specify co-translational degradation of the mRNA for β-tubulin, but not for α-tubulin [47]. Thus, it is likely that a combination of differential mRNA stability and "translate-ability" contributes to the overall β- and α-tubulin abundance in the different developmental stages of trypanosomes, an area that merits further study. ESTs encoding proteins known to be associated with the flagellum are about half as abundant in the PCF (1%) as in the BSF (1.9%) (Fig. 4).

3.4.3. Cysteine proteases of T. congolense

Cysteine proteases in parasitic organisms and their inhibitors have attracted considerable interest [48], none perhaps more so than in *T. congolense* (reviewed in [49]). Proteases are thought to play direct roles in disease pathogenesis as virulence factors involved in host invasion, migration, metabolism and immune evasion. African cattle breeds that are 'trypanotolerant', *i.e.*, that are able to partially control a *T. congolense* infection and limit its associated pathology, have been found to develop a much more robust IgG antibody response against a *T. congolense* cysteine protease (a cathepsin L-like protease called congopain) compared to 'trypanosensitive' breeds [49]. This observation has led to the suggestion that trypanotolerant cattle may control the disease by antibody-mediated neutralization of *T. congolense* cysteine proteases and that immunization against these cysteine proteases might be effective in minimizing the disease in cattle, if not the parasite itself [50]. Recently, a new family of cathepsin B-like cysteine proteases in *T. congolense* has also been identified that are encoded by at least 13 non-tandem genes in the genome, indicating that the abundance, diversity, and possible biological roles of cysteine proteases in *T. congolense* are much greater than previously suspected [51].

The presence of ESTs encoding cathepsin-like cysteine proteases is consistent with this possibility. After kinases and phosphatases (Fig. 3B), more ESTs for cysteine proteases were found in each of the four libraries than for any other single enzyme class (Fig. 1). The percent of the cysteine protease ESTs is slightly higher in the MCF/BSF libraries (1.3% and 1.0%, respectively) than in the PCF/EMF libraries (0.5% and 0.8%), which may assist MCF cells in their confrontation with many defense factors immediately upon their entry in the mammalian host. ESTs encoding cysteine protease inhibitors were also relatively abundant (0.5% of the MCF library; Fig. 1), suggesting
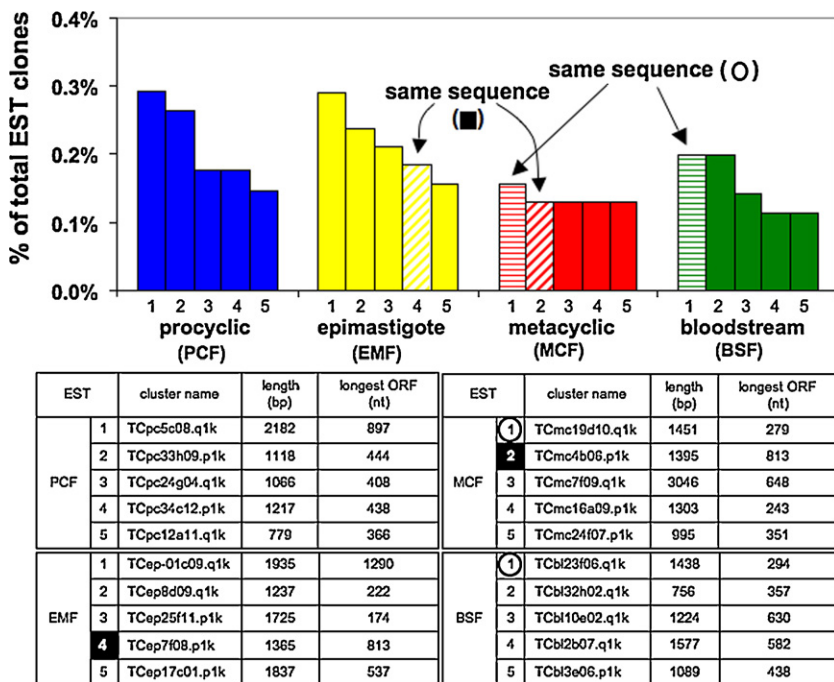
Fig. 6. Relative abundances of the five most abundant "hypothetical" EST clusters in each of the indicated libraries, and their names and properties. The black square and open circle denote the same EST sequence in two different libraries.

that regulation of cysteine protease activity is also important during the parasite's interactions with its mammalian hosts and insect vector.

3.5. The five most abundant ESTs encoding proteins of unknown function in each of developmental stage

About one-fourth (21–28%) of the *T. congolense* ESTs in each of the four non-normalized libraries have homology to genes of other organisms encoding hypothetical proteins of unknown function (Table 1). In almost all cases, the greatest homology is with predicted *T. brucei* and *T. cruzi* proteins of unknown function. The five most abundant hypothetical EST clusters in each of these libraries are summarized in Fig. 6. None of these hypothetical ESTs is particularly abundant, ranging between 0.1% and 0.3% of the total ESTs in their respective library. In two cases, a "top-five" hypothetical EST occurs in two of the four non-normalized libraries.

The most abundant "hypothetical" EST cluster in both the MCF and BSF libraries is nearly identical (indicated by the open circle in Fig. 6), an approximately 1450-nt sequence bearing a 5′ spliced leader and 3′ poly(A). The longest open translation reading frame (ORF) within this sequence is 279 nts in the MCF EST cluster (TCmc19d10.q1k) and 294 nts in the BSF EST cluster (Tcbl23f06.q1k), a difference caused by a single nucleotide insertion near the end of the BSF ORF that extends its length beyond that of the MCF ORF. Both EST clusters have a 3′ UTR of more than 1000 nts. The corresponding protein of 93 aa (MCF) or 98 aa (BSF) has extensive homology with a hypothetical protein encoded by the *T. brucei* genome and weak homology (e^{-11}) with a "heat shock factor binding protein" whose gene has been previously detected in the genomes of organisms ranging from humans to *Drosophila*. Studies in *Caenorhabditis elegans* suggest this protein is a negative regulator of the heat shock response [52]. Since MCF cells experience a sudden temperature increase when they enter a mammalian host, the presence of this protein could downregulate a potential deleterious heat shock response associated with the MCF-to-BSF transition. Consistent with this possibil-

ity, ESTs for this protein were not detected in the PCF and EMF libraries.

The other shared "top-five" hypothetical EST was detected in the EMF and MCF libraries (indicated by the black square in Fig. 6). This EST of about 1.4 kb encodes a protein of 271 aa with homology to a potential *T. brucei* protein (e^{-28}), but no substantive similarity to deduced proteins encoded by genomes of non-trypanosomatid organisms. This EST was not detected in the PCF and BSF libraries. Thus, its deduced protein may play a role unique to the epimastigote and metacyclic stages, the two stages of the African trypanosome life cycle about which the least is known.

The remaining "top-five" hypothetical ESTs have little in common with each other, other than their similarity to genes encoding proteins of unknown function in the *T. brucei* and *T. cruzi* genomes. The most abundant hypothetical EST (TCep-01c09.q1k) in the EMF library is homologous to a multi-copy, tandem gene family in *T. brucei*, whereas the "top-five" ESTs in the other libraries are typically homologous to single-copy *T. brucei* genes.

3.6. The five most abundant "no hit" ESTs in each developmental stage

Between 27% (PCF) and 45% (MCF) of the ESTs in the libraries have an alignment parameter of 50 or less with known sequences in other organisms, including other trypanosomatids, *i.e.*, they do not have a significant match with any other sequence (Table 1). The five most abundant "no hit" EST clusters in each library are summarized in Fig. 7. Since these "no hit" ESTs are not similar to sequences in the *T. brucei* and *T. cruzi* genomes at either the nucleotide level or the deduced amino acid sequence level, no hints about their possible functions are available. However, at least some of these ESTs are likely to encode *T. congolense*-specific proteins and/or non-protein functions. For example, the fifth most abundant "no hit" EST cluster in the BSF library (TCbl36f12.q1k) has a coding sequence of 1191 nts, encoding a putative *T. congolense*-specific protein of nearly 400 aa with a predicted signal peptide. Some, but not all, of the other "no hit" ESTs have 5′ spliced leader and/or 3′ poly(A) sequences, indicating they are derived from a polycistronic transcript, as are
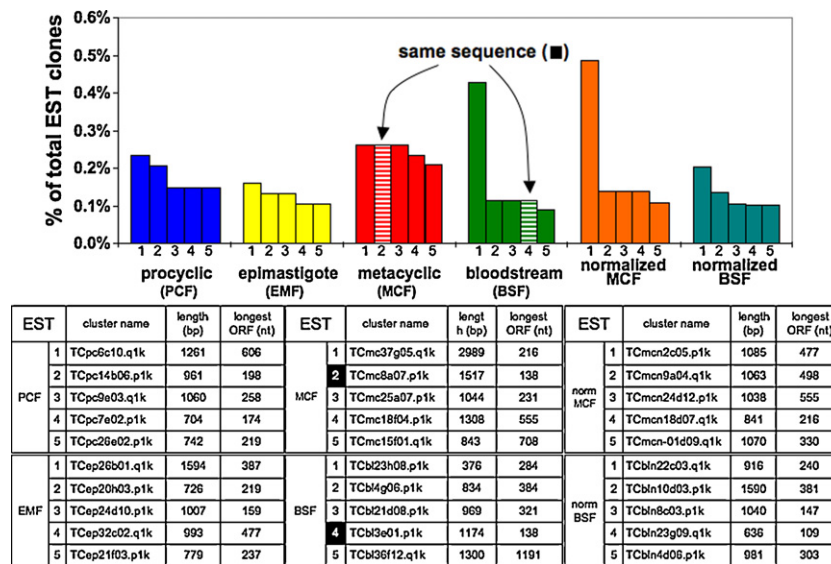
EST		cluster name	length (bp)	longest ORF (nt)	EST		cluster name	length (bp)	longest ORF (nt)	EST		cluster name	length (bp)	longest ORF (nt)
PCF	1	TCpc6c10.q1k	1261	606	MCF	1	TCmc37g05.q1k	2989	216	norm MCF	1	TCmcn2c05.p1k	1085	477
	2	TCpc14b06.p1k	961	198		2	TCmc8a07.p1k	1517	138		2	TCmcn9a04.q1k	1063	498
	3	TCpc9e03.q1k	1060	258		3	TCmc25a07.p1k	1044	231		3	TCmcn24d12.p1k	1038	555
	4	TCpc7e02.p1k	704	174		4	TCmc18f04.p1k	1308	555		4	TCmcn18d07.q1k	841	216
	5	TCpc26e02.p1k	742	219		5	TCmc15f01.q1k	843	708		5	TCmcn-01d09.q1k	1070	330
EMF	1	TCep26b01.q1k	1594	387	BSF	1	TCbl23h08.p1k	376	284	norm BSF	1	TCbln22c03.q1k	916	240
	2	TCep20h03.p1k	726	219		2	TCbl4g06.p1k	834	384		2	TCbln10d03.p1k	1590	381
	3	TCep24d10.p1k	1007	159		3	TCbl21d08.p1k	969	321		3	TCbln8c03.p1k	1040	147
	4	TCep32c02.q1k	993	477		4	TCbl3e01.p1k	1174	138		4	TCbln23g09.q1k	636	109
	5	TCep21f03.p1k	779	237		5	TCbl36f12.q1k	1300	1191		5	TCbln4d06.p1k	981	303

Fig. 7. Relative abundances of the five most abundant "no hit" EST clusters in each of the indicated libraries, and their names and properties. The black square denotes the same EST sequence in two different libraries.

probably all mature mRNAs in African trypanosomes [11]. Others have no obvious distinguishing features, suggesting they may be primarily derived from transcribed intergenic regions or untranslated regions.

Only one of the top five "no hit" EST clusters is shared by two libraries (MCF and BSF; indicated by the black square in Fig. 7). This MCF EST cluster (TCmc8a07.p1k) is longer at one end than the corresponding BSF EST cluster (TCbl3e01.p1k), accounting for the difference in length (1517 bp versus 1174 bp, respectively). The longest ORF in both of these EST clusters is a short 138 nts, potentially encoding a protein of only 46 aa. Neither EST cluster has a 5′ spliced leader or 3′ poly(A) and they are only present in the MCF and BSF libraries, suggesting stage-specific regulation. In considering all of these "no hit" ESTs it is worth noting that ORF-less or short ORF RNAs could be produced from a significant proportion of all trypanosome intergenic regions and might be functionless by-products of cis-splicing and polyadenylation [53–55]. Alternatively, these RNAs could contribute important non-protein regulatory roles yet unknown, a possibility attracting substantial interest in many other eukaryotic organisms [56].

The most abundant of all of the top five "no hit" ESTs in the four non-normalized libraries occurs in the BSF library (TCbl23h08.p1k) and constitutes 0.43% of the BSF ESTs (15 of 3523 clones). This relative short EST cluster (376 bp) has a run of 21 T's near one end and a run of 16 A's near the other end, and might be a cDNA synthesis/cloning artifact. It does not possess a spliced leader sequence.

3.7. Rare transcripts

Two of the four regular cDNA libraries (MCF and BSF) were subjected to a normalization procedure (see Materials and Methods) designed to reduce abundant ESTs in the library [28,29]. The normalization of the MCF library resulted in a decrease of the VSG ESTs from 5.8% of the library to about 0.4%, whereas normalization of the BSF library resulted in a corresponding drop of VSG ESTs from 5.2% to 0.9% (not shown). Likewise, the normalization caused the β- and α-tubulin ESTs in both libraries to drop 5–10-fold, as well as most of the ribosomal protein ESTs (not shown). After normalization, 75% of the ESTs in the normalized (norm)MCF library were "hypothetical" or "no hit" ESTs, and 70% of the normBSF ESTs were in the same categories (Table 1).

Since normalization selects for the less abundant sequences, we also examined the five most abundant "no hit" ESTs in the normMCF and normBSF libraries, in anticipation that normalization would enrich for rare *T. congolense*-specific RNA sequences in these two developmental stages (Fig. 7). Indeed, none of the top five "no hit" ESTs in either normalized library is the same as the top five "no hit" ESTs in the corresponding non-normalized libraries. Likewise, the two normalized libraries do not share any of their top five "no hit" EST clusters. In the normMCF library the most abundant of all the EST clusters is a "no hit" EST (Tmcn2c05.p1k) that constituted 0.48% of the clones sequenced in this library (14 of 2892 clones). This 1085-bp sequence has an ORF of 477 nts, but has numerous internal runs of all four nucleotides that are seven residues or more in length, suggesting but not proving that it is an intergenic sequence. Likewise, several of the other normalized top five "no hit" ESTs have internal runs of nucleotides suggesting they may not be derived from the coding portion of genes. Other normalized top five "no hit" ESTs, however, have a 5′ spliced leader spliced leader addition and 3′ polyadenylation that accompanies generation of mature trypanosome mRNAs.

In summary, this analysis of ESTs from the four main developmental stages of *T. congolense* is consistent with and confirms many of the biological, metabolic, immunological and molecular features reported for *T. congolense* and African trypanosomes in general over the past four decades. It has also provided a foundation for a more detailed examination of these properties and revealed many new stage-specific genes/gene products, in particular those of the epimastigote and metacyclic forms, which can be much more readily cultured and studied with *T. congolense* than with other African trypanosome species.

References

[1] Hoare CA. The trypanosomes of mammals. Oxford/Edinburgh: Blackwell Scientific Publications; 1972.

[2] Pays E, Vanhollebeke B. Mutual self-defence: the trypanolytic factor story. Microbes Infect 2008;10:985–9 [review].

[3] Widener J, Nielsen MJ, Shiflett A, Moestrup SK, Hajduk S. Hemoglobin is a cofactor of human trypanosome lytic factor. PLoS Pathog 2007;3:1250–61.

[4] Berriman M, Ghedin E, Hertz-Fowler C, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309:416–22.

[5] Broden A. Les infections à trypanosomes au Congo chez l'homme et les animaux (communication préliminaire). Bull Soc D'Etudes Coloniales (Brussels), February 1904:116–39.

[6] Shaw PMA. Economics of African trypanosomiasis. In: Maudlin I, Holmes PH, Miles MA, editors. Trypanosomiases. Wallingord, UK: CABI Publishing; 2004. p. 369–402.

[7] Bengaly Z, Sidibe I, Ganaba R, Desquesnes M, Boly H, Sawadogo L. Comparative pathogenicity of three genetically distinct types of *Trypanosoma congolense* in cattle: clinical observations and haematological changes. Vet Parasitol 2002;108:1–19.

[8] Bienen EJ, Webster P, Fish WR. *Trypanosoma (Nannomonas) congolense*: changes in respiratory metabolism during the life cycle. Exp Parasitol 1991;73:403–12.

[9] Vickerman K. Developmental cycles and biology of pathogenic trypanosomes. Br Med Bull 1985;41:105–14.

[10] Turner MJ. The biochemistry of the surface antigens of the African trypanosomes. Br Med Bull 1985;41:137–43.

[11] Donelson JE. Antigenic variation and the African trypanosome genome. Acta Trop 2003;85:391–404 [review].

[12] Roditi I, Lehane MJ. Interactions between trypanosomes and tsetse flies. Curr Opin Microbiol 2008;11:345–51 [review].

[13] Maudlin I, Welburn SC. Maturation of trypanosome infections in tsetse. Exp Parasitol 1994;79:202–5.

[14] Utz S, Roditi I, Renggli CK, Almeida IC, Acosta-Serrano A, Bütikofer P. *Trypanosoma congolense* procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. Eukaryot Cell 2006;5:1430–40.

[15] Bayne RA, Kilbride EA, Lainson FA, Tetley L, Barry JD. A major surface antigen of procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol 1993;61:295–310.

[16] Beecroft RP, Roditi I, Pearson TW. Identification and characterization of an acidic major surface glycoprotein from procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol 1993;61:285–94.

[17] Asbeck K, Ruepp S, Roditi I, Gibson W. GARP is highly conserved among *Trypanosoma congolense* Savannah, Forest and Kilifi subgroups. Mol Biochem Parasitol 2000;106:303–6.

[18] Sakurai T, Sugimoto C, Inoue N. Identification and molecular characterization of a novel stage-specific surface protein of *Trypanosoma congolense* epimastigotes. Mol Biochem Parasitol 2008;161:1–11.

[19] Vassella E, Oberle M, Urwyler S, et al. Major surface glycoproteins of insect forms of *Trypanosoma brucei* are not essential for cyclical transmission by tsetse. PLoS ONE 2009;4:e4493 [Epub February 18, 2009].

[20] Lenardo MJ, Rice-Ficht AC, Kelly G, Esser KM, Donelson JE. Characterization of the genes specifying two metacyclic variable antigen types in *Trypanosoma brucei rhodesiense*. Proc Natl Acad Sci USA 1984;81:6642–6.

[21] Crowe JS, Barry JD, Luckins AG, Ross CA, Vickerman K. All metacyclic variable antigen types of *Trypanosoma congolense* identified using monoclonal antibodies. Nature 1983;306:389–91.

[22] Hajduk S, Vickerman K. Antigenic differentiation of *Trypanosoma brucei*: studies on metacyclic and first parasitaemia populations. Trans R Soc Trop Med Hyg 1981;75:145–6.

[23] Hirumi H, Hirumi K. *In vitro* cultivation of *Trypanosoma congolense* bloodstream forms in the absence of feeder cell layers. Parasitology 1991;102(Suppl 2):225–36.

[24] Ross CA. *Trypanosoma congolense*: differentiation to metacyclic trypanosomes in culture depends on the concentration of glutamine or proline. Acta Trop 1987;44:293–301.

[25] Lanham SM, Godfrey DG. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. Exp Parasitol 1970;28:521–34.

[26] Inoue N, Lluz AT, Mori T, Nagasawa H, Fujisak K, Mikami T. Novel species specific antigens of *Trypanosoma congolense* and their different localization among life-cycle stages. J Vet Med Sci 2000;62:1041–5.

[27] Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nat Prod 2006;1:581–5.

[28] Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 1994;91:9228–32.

[29] Bonaldo MF, Lennon G, Soares MB. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 1996;6:791–806.

[30] Bonfield JK, Smith K, Staden R. A new DNA sequence assembly program. Nucl Acids Res 1995;23:4992–9.

[31] The Universal Protein Resource (UniProt) 2009. UniProt Consortium. Nucleic Acids Res 2009;37(January(Database issue)):D169–74.

[32] Berriman M, Rutherford K. Viewing and annotating sequence data with Artemis. Brief Bioinform 2003;4:124–32.

[33] Chenna R, Sugawara H, Koike T, et al. Multiple sequence alignment with the Clustal series of programs. Nucl Acids Res 2003;31:3497–500.

[34] Bütikofer P, Vassella E, Boschung M, et al. Glycosylphosphatidylinositol-anchored surface molecules of *Trypanosoma congolense* insect forms are developmentally regulated in the tsetse fly. Mol Biochem Parasitol 2002;119:7–16.

[35] Eshita Y, Urakawa T, Hirumi H, Fish WR, Majiwa PA. Metacyclic form-specific variable surface glycoprotein-encoding genes of *Trypanosoma (Nannomonas) congolense*. Gene 1992;113:139–48.

[36] Rausch S, Shayan P, Salnikoff J, Reinwald E. Sequence determination of three variable surface glycoproteins from *Trypanosoma congolense*. Conserved sequence and structural motifs. Eur J Biochem 1994;223:813–21.

[37] Urakawa T, Eshita Y, Majiwa PA. The primary structure of *Trypanosoma (Nannomonas) congolense* variant surface glycoproteins. Exp Parasitol 1997;85:215–24.

[38] Esser KM, Schoenbechler MJ, Gingrich JB. *Trypanosoma rhodesiense* blood forms express all antigen specificities relevant to protection against metacyclic (insect form) challenge. J Immunol 1982;129:1715–8.

[39] Turner CM, Barry JD, Maudlin I, Vickerman K. An estimate of the size of the metacyclic variable antigen repertoire of *Trypanosoma brucei rhodesiense*. Parasitology 1988;97:269–76.

[40] El-Sayed NM, Myler PJ, Blandin G, et al. Comparative genomics of trypanosomatid parasitic protozoa. Science 2005;309:404–9.

[41] El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005;309:409–15.

[42] Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. Science 2005;309:436–42.

[43] van Hellemond JJ, Opperdoes FR, Tielens AG. The extraordinary mitochondrion and unusual citric acid cycle in *Trypanosoma brucei*. Biochem Soc Trans 2005;33:967–71 [review].

[44] Bringaud F, Rivière L, Coustou V. Energy metabolism of trypanosomatids: adaptation to available carbon sources. Mol Biochem Parasitol 2006;149:1–9 [review].

[45] van Weelden SW, Fast B, Vogt A, et al. Procyclic *Trypanosoma brucei* do not use Krebs cycle activity for energy generation. J Biol Chem 2003;278:12854–63.

[46] Bryan J, Wilson L. Are cytoplasmic microtubules heteropolymers? Proc Natl Acad Sci USA 1971;8:1762–6.

[47] Bachurski CJ, Theodorakis NG, Coulson RM, Cleveland DW. An amino-terminal tetrapeptide specifies cotranslational degradation of beta-tubulin but not alpha-tubulin mRNAs. Mol Cell Biol 1994;14:4076–86.

[48] McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M. Proteases in parasitic diseases. Annu Rev Pathol 2006;1:497–536.

[49] Lalmanach G, Boulangé A, Serveau C, et al. Congopain from *Trypanosoma congolense*: drug target and vaccine candidate. Biol Chem 2002;383:739–49.

[50] Authié E, Boulangé A, Muteti D, Lalmanach G, Gauthier F, Musoke AJ. Immunisation of cattle with cysteine proteinases of *Trypanosoma congolense*: targetting the disease rather than the parasite. Int J Parasitol 2001;31:1429–33.

[51] Mendoza-Palomares C, Biteau N, Giroud C, et al. Molecular and biochemical characterization of a cathepsin B-like protease family unique to *Trypanosoma congolense*. Eukaryot Cell 2008;7:684–97.

[52] Satyal SH, Chen D, Fox SG, Kramer JM, Morimoto RI. Negative regulation of the heat shock transcriptional response by HSBP1. Genes Dev 1998;12:1962–74.

[53] Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL. Messenger RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol 2005;143:125–34.

[54] Hug M, Hotz HR, Hartmann C, Clayton C. Hierarchies of RNA-processing signals in a trypanosome surface antigen mRNA precursor. Mol Cell Biol 1994;14:7428–35.

[55] Schürch N, Hehl A, Vassella E, Braun R, Roditi I. Accurate polyadenylation of procyclin mRNAs in *Trypanosoma brucei* is determined by pyrimidine-rich elements in the intergenic regions. Mol Cell Biol 1994;14:3668–75.

[56] Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF. RNA regulation of epigenetic processes. BioEssays 2009;31:51–9.